Administrative

- HW3 grades will be available this Sunday
  - Discussions will be posted on Piazza
- HW4 released
  - Due Nov 14
- Project deliverables I due Monday, Oct 31
  - 1-2 page report on what has been done so far, what else remains to do
    - Due 11:59pm, Oct 31 EST
  - In class presentation: 5 min to talk + 2 min Q&A
    - Due **3pm**, Oct 31 EST
  - 14 groups in total, so please keep to time limits

18734 Recitation

Review of Statistical distance

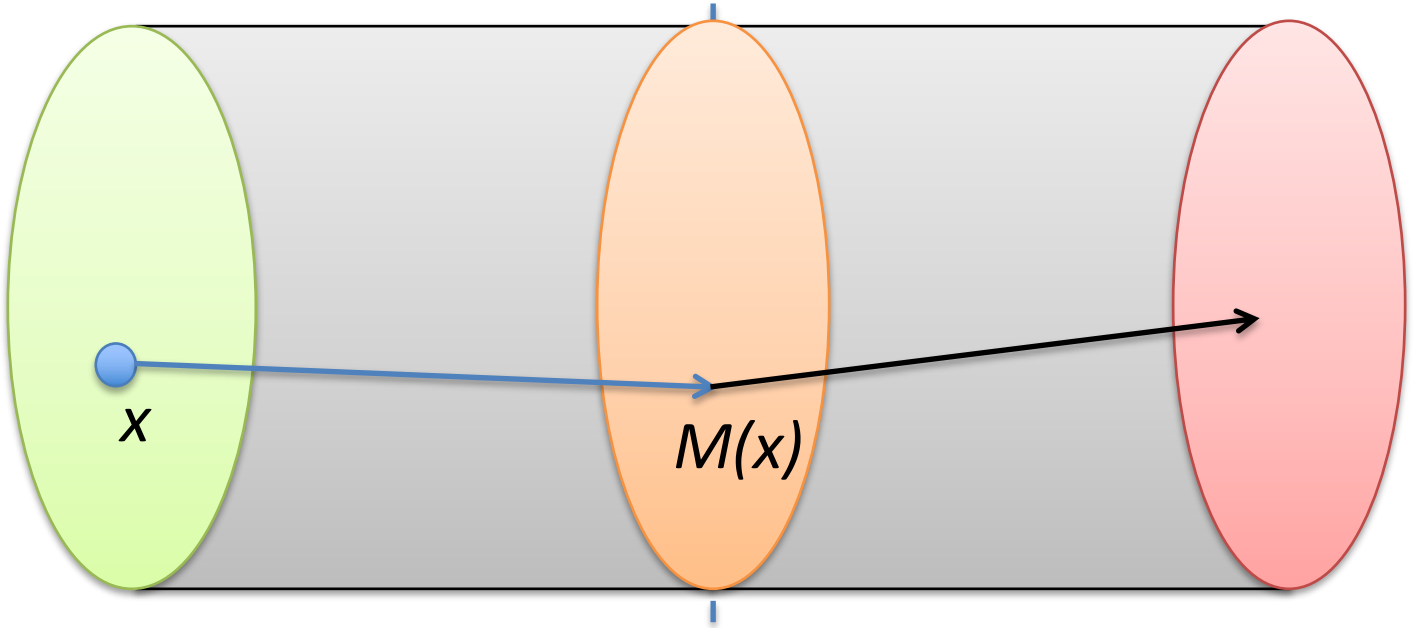
Individual fairness

Classifier
(eg. ad network)

Vendor
(eg. capital one)

M: V → O

F: O → A



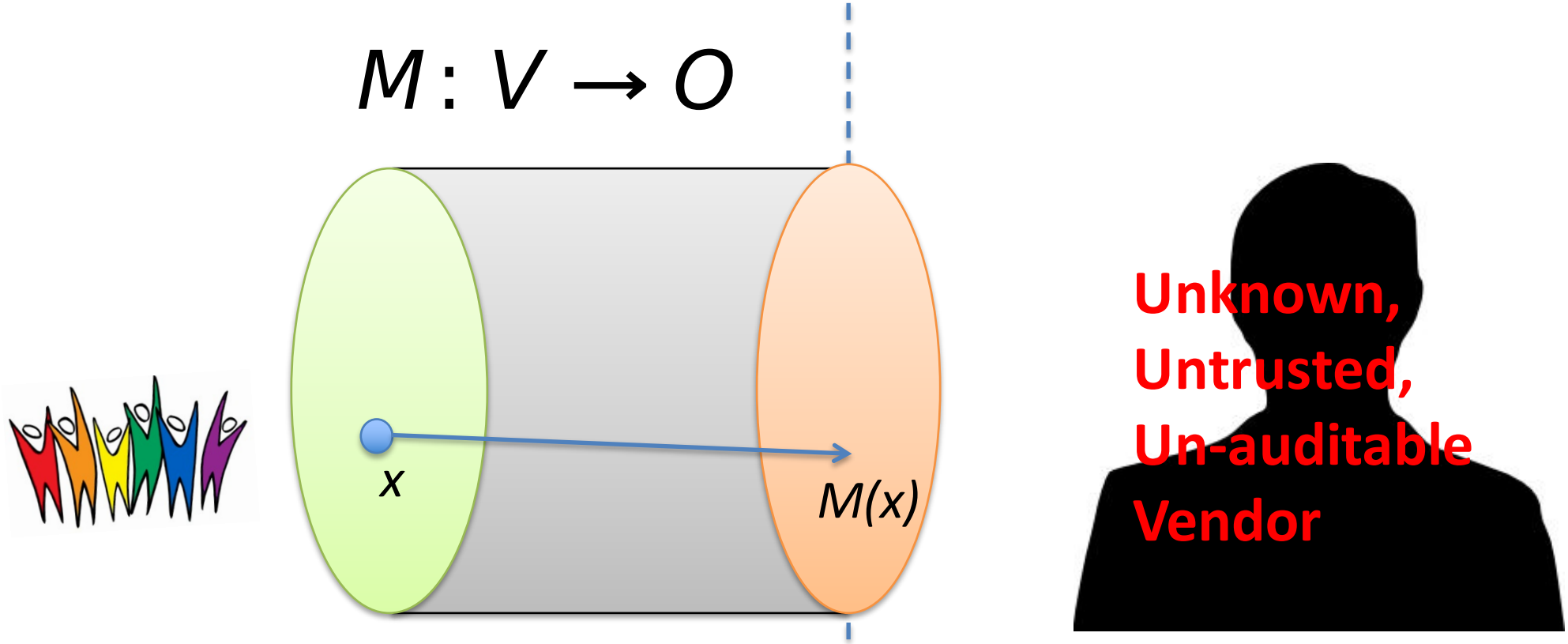
V: Individuals

O: outcomes

A: actions

Goal:

Achieve Fairness in the classification step



The similarity metric

- The extent to which pairs of individuals should be regarded as similar for classification
- Expresses ground truth/ best approximation according to societal norms
- Example of a metric:
  - People with a credit scores of 310 are similar to those with credit scores of 300 and should be ***not*** be allowed to get a loan

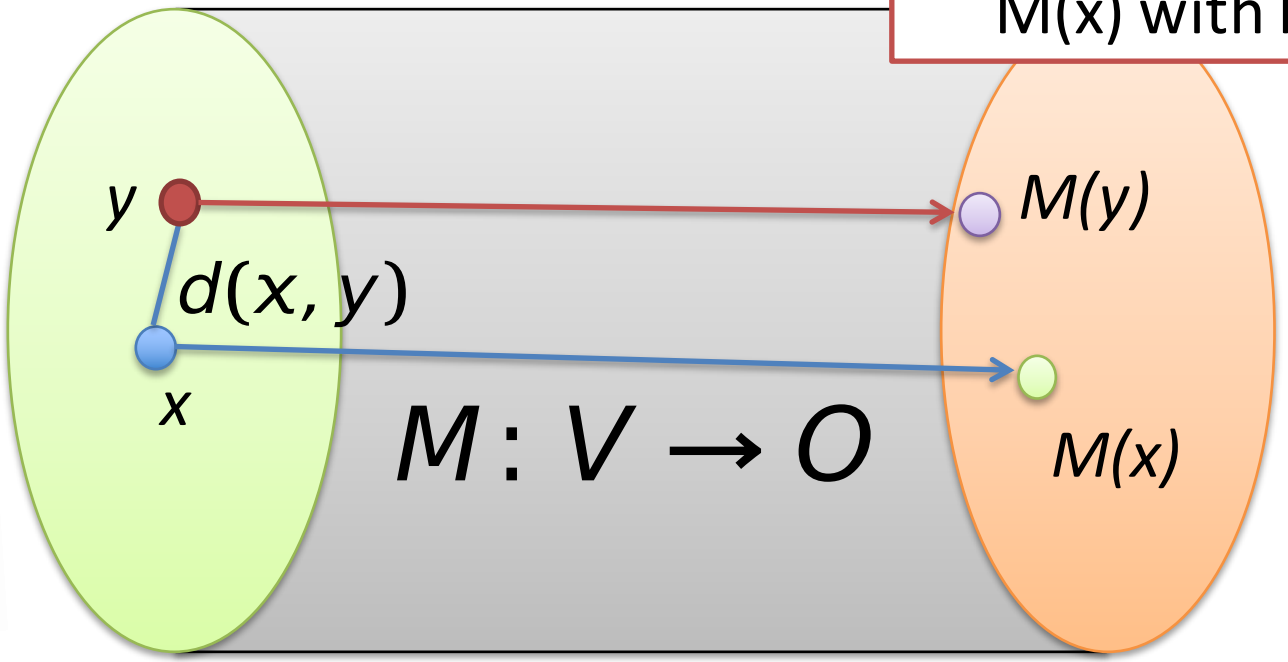
Constructing a similarity metric for a classifier (e.g. ad network)

Formalizing similarity between individuals and outcomes

Think of V as space with metric d(x,y)

similar = small d(x,y)

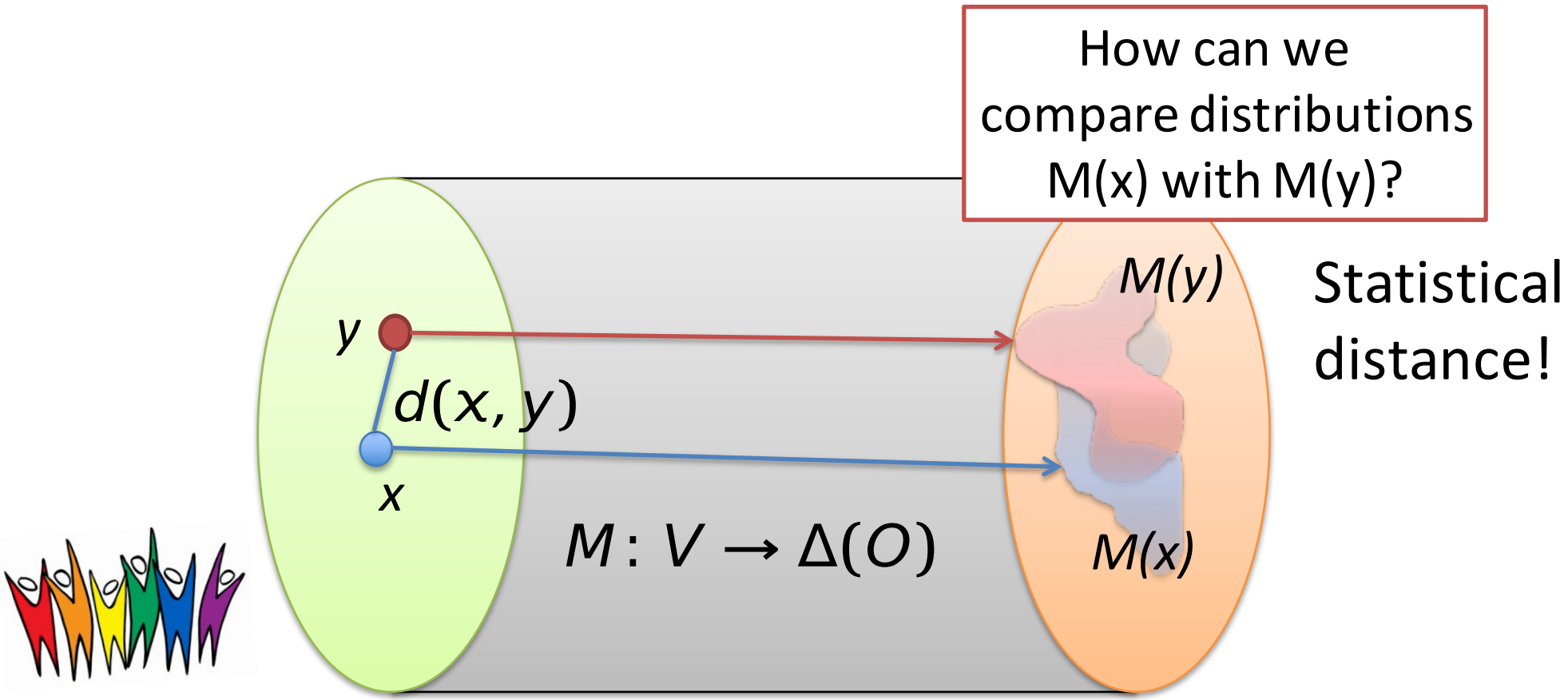
How can we compare distributions M(x) with M(y)?



V: Individuals

O: outcomes

Distributional outcomes



V: Individuals

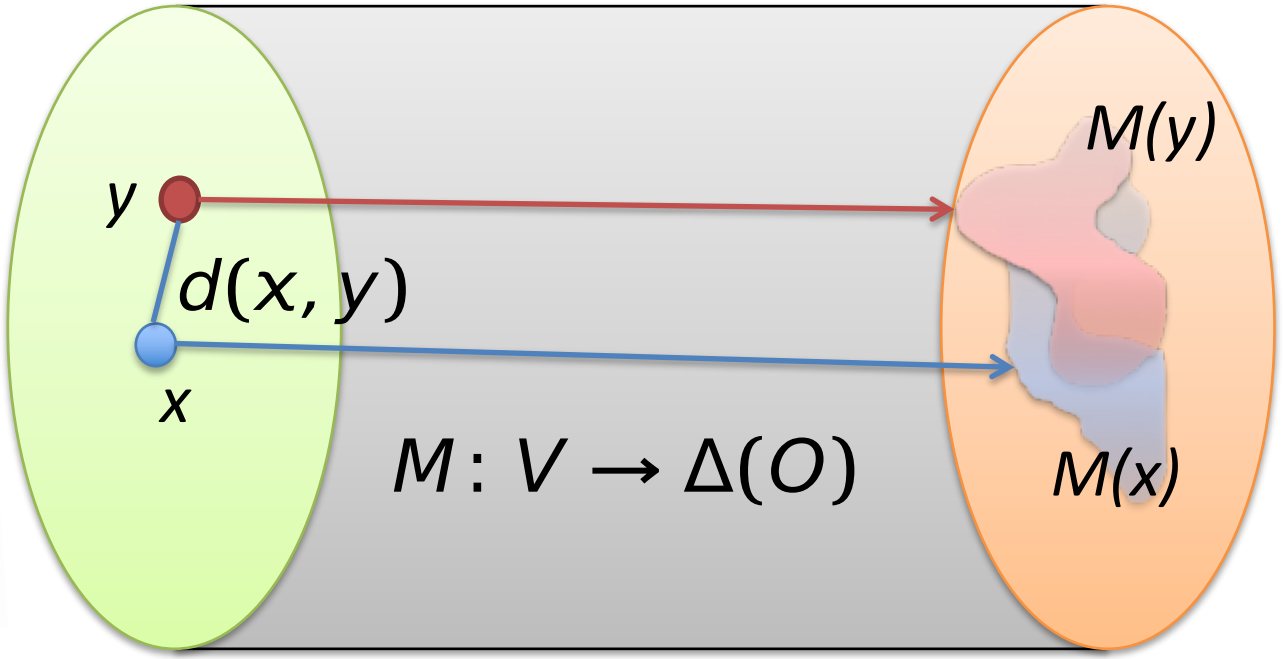
O: outcomes

Metric $d : V \times V \to \mathbb{R}$

Lipschitz condition $\|M(x) - M(y)\| \leq d(x, y)$

Statistical distance between distributions

in [0,1]



V: Individuals

O: outcomes

Statistical Distance

P, Q denote probability measures on a finite domain A. The *statistical distance* between P and Q is denoted by

$$D_{\mathrm{tv}}(P, Q) = \frac{1}{2} \sum_{a \in A} |P(a) - Q(a)| .$$

Example: High D

A= {0,1}

P(0) = 1, P(1) = 0

Q(0) = 0, Q(1) = 1

D(P, Q) = 1

Statistical Distance

P, Q denote probability measures on a finite domain A. The *statistical distance* between P and Q is denoted by

$$D_{\mathrm{tv}}(P, Q) = \frac{1}{2} \sum_{a \in A} |P(a) - Q(a)| .$$

Example: Low D

A= {0,1}

P(0) = 1, P(1) = 0

Q(0) = 1, Q(1) = 0

D(P, Q) = 0

Statistical Distance

P, Q denote probability measures on a finite domain A. The *statistical distance* between P and Q is denoted by

$$D_{\mathrm{tv}}(P, Q) = \frac{1}{2} \sum_{a \in A} |P(a) - Q(a)| .$$

Example: Mid D

A= {0,1}

P(0) = P(1) = ½

Q(0) = ¾, Q(1) = ¼

D(P, Q) = ¼